



1. はじめに
2. 金融テキストマイニング研究
3. テキストデータによる長期的な市場分析手法
4. 運用テストによる他手法との比較

経済テキスト情報を用いた 長期的な市場動向推定

Long-term Financial Market Analysis Using Economic Textual
Information

Itaru Aso

Graduate School of Information Engineering, Toyama Prefectural University
t855001@st.pu-toyama.ac.jp

L205, AM 9:00-9:25, Friday, December 8, 2018,
Toyama Prefectural Univ.



1.1 背景

本研究の背景

- 1 テキスト情報と市場変動の関係性を市場分析に応用する研究が増えている
- 2 経済指標やテクニカル指標などの数値情報には指標化されない情報をテキスト情報から抽出することが期待されている

既存研究

- 1 twitter のテキストデータから翌日のダウ・ジョーンズ工業株価平均の変動の予測
- 2 個別銘柄に関する掲示板データから翌日の株価リターンや出来高の予測



1.3 本研究の目的

本研究の目的

数週間以上の長期的でかつ個別銘柄よりも広範な市場予測を行う

本研究の概要

- 1 テキストマイニングにより経済テキスト情報を分析
- 2 経済指標分析や時系列データ分析などの既存研究と運用テストや変動予測精度の比較
- 3 テキストマイニング手法を長期的な市場分析に用いることの優位性を示す

1. はじめに
2. 金融テキスト
マイニング研究

3. テキストデータ
による長期的な市
場分析手法

4. 運用テストによ
る他手法との比較



2.1 概観

- はじめに
- 金融テキストマイニング研究
- テキストデータによる長期的な市場分析手法
- 運用テストによる他手法との比較

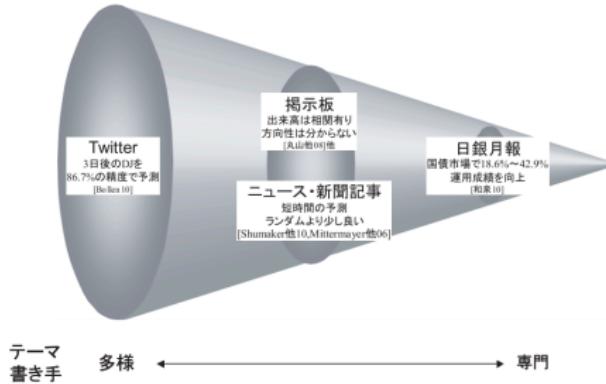


Figure: 1. 金融テキストマイニング研究の概観

概観

- 1 twitter やブログが 1 番多様な内容と書き手をもつテキスト
- 2 掲示板や Web 上の経済ニュースのテキストは、金融機関のトレーダーも参考にしている
- 3 経済の専門家たちが他の専門家や投資家に市場の状況を解説するため書いたテキスト



3.1 本研究の分析手法

本研究では長期的な市場分析に有効なテキストデータとして、日本銀行の金融経済月報を用いた。

日本銀行の金融経済月報

- ① 毎月半ばに A4 で 15-20 ページにわたり公開されている
- ② 日本銀行の経済状況に対する態度が含まれており、金融市場のトレーダーが着目している
- ③ 一貫性のある形式であり、異なる時点間のテキスト内容の変化が比較しやすい

分析手法

- ① 時間的な出現パターンから単語のグループ化
- ② 主成分分析による次元削減
- ③ 回帰分析による予測



3.2 共起関係に基づく主要単語の抽出

共起関係の抽出

- はじめに
- 金融テキストマイニング研究
- テキストデータによる長期的な市場分析手法
- 運用テストによる他手法との比較

各月 t のテキストデータ $D(t)$ の特徴を表す主要単語を抽出する

- 高頻出単語 : $D(t)$ 中の名詞・動詞・形容詞のうち, 頻度が上位 M 個高頻度語の集合 $HighFreq = w_i, i = 1, \dots, M$ を抽出する
- $HighFreq$ 内の単語 w_i, w_j 間の共起度 $co(w_i, w_j)$ を Jaccard 係数で計算し, 上位の L 個の単語間をリンクで結ぶ

$$co(w_i, w_j) = \frac{D(t) \text{ のうち } w_i, w_j \text{ が共に出現する段落数}}{w_i, w_j \text{ のうち少なくとも一方が出現する段落数}} \quad (1)$$

- $D(t)$ 中のすべての語 w に対して, すべての土台 g が考慮されたときに語 w が用いられる条件付き確率 $key(w)$ を計算して, 上位 N 個を土台間を結ぶキーワード $HighKey$ として主要単語に加える

$$key(w) = \left(1 - \prod_g \left(1 - \frac{w \text{ と } g \text{ 中の語の共起度}}{g \text{ 中の語の出現頻度}} \right) \right) \quad (2)$$



3.3 主成分分析による単語のグループ化

- はじめに
- 金融テキストマイニング研究
- テキストデータによる長期的な市場分析手法
- 運用テストによる他手法との比較

特徴量抽出

過去の一定期間 $t_{-1} \cdots t_{-1}$ の各月 t の主要単語 $HighFreq(t)$ と $HighKey(t)$ が含まれる語の出現パターンに対して主成分分析を行い、特徴量の次元圧縮を行う。

$$A(w_i, t) = \begin{cases} 1 & w_i \in \{HighFreq(t), HighKey(t)\} \\ 0 & (otherwise) \end{cases} \quad (3)$$

この行列に対して、主成分分析により N_{pc} 個の合成成分にまとめる各月の N_{pc} 個の主成分スコアを対象期間について時系列順に並べることによって、 N_{pc} 次元の時系列データ $x_i(t)$, $i = \{1 \cdots N_{pc}\}$ が作成される



3.4 回帰分析による市場データの動向分析

回帰分析

- はじめに
- 金融テキストマイニング研究
- テキストデータによる長期的な市場分析手法
- 運用テストによる他手法との比較

各主成分スコアの各月の動き $x_i(t)$ から月次での価格の動きを解析する。

$x_i(t)$ を説明変数とし、各月の月末の価格データ $p(t)$ を目的変数として重回帰分析を行う

$$\bar{p}(t) = a_0 + \sum_{i=1}^{N_{pc}} a_i x_i(t) \quad (4)$$



4.1 運用テストの手法

運用テスト

1. はじめに
2. 金融テキスト
マイニング研究
3. テキストデータ
による長期的な市
場分析手法
4. 運用テストによ
る他手法との比較

- 1 環境は、日本国債市場
- 2 売買は月次
- 3 取引量は毎月決まった資本量とし、固定
- 4 取引手数料は考慮しない
- 5 売買ルールによって予測された変動幅を



3.7 運用テスト結果

- はじめに
- 金融テキスト
マイニング研究
- テキストデータ
による長期的な市
場分析手法
- 運用テストによ
る他手法との比較

	CPR	TF-SVR	Co-SVR	BOJ	EXT
日本国債 2 年	59.60	40.85	36.49	50.24	41.70
日本国債 5 年	223.18	88.56	215.37	-39.90	143.70
日本国債 10 年	243.02	248.14	233.70	31.79	-47.84

単位はベースポイント (0.01%)。太字は各市場での最大利益。

Figure: 2. 平均損益 (年率)

	CPR	TF-SVR	Co-SVR	BOJ	EXT
日本国債 2 年	55.17	65.52	58.62	62.07	55.17
日本国債 5 年	58.62	44.83	58.62	55.17	62.07
日本国債 10 年	55.17	62.07	58.62	55.17	44.83

Figure: 3. テスト期間の正答率

	CPR	TF-SVR	Co-SVR	BOJ	EXT
日本国債 2 年	85.71	71.43	71.43	71.43	57.14
日本国債 5 年	85.71	57.14	42.86	28.57	42.86
日本国債 10 年	71.43	71.43	85.71	42.86	57.14

Figure: 4. 高変動期の正答率



4.1 まとめ

運用テスト

- はじめに
- 金融テキストマイニング研究
- テキストデータによる長期的な市場分析手法
- 運用テストによる他手法との比較

- ① 先行研究でよく使われる単語の頻度をそのままサポートベクトル回帰により予測する手法より精度は向上した
- ② 金融経済月報に記述されている日本銀行の態度変化に関する兆候や意図をうまく抽出できた